

МИНОБРНАУКИ РОССИИ



**Федеральное государственное бюджетное образовательное учреждение
высшего образования
"Российский государственный гуманитарный университет"
(ФГБОУ ВО "РГГУ")**

**ИНСТИТУТ ИНФОРМАЦИОННЫХ НАУК И ТЕХНОЛОГИЙ БЕЗОПАСНОСТИ
ФАКУЛЬТЕТ ИНФОРМАЦИОННЫХ СИСТЕМ И БЕЗОПАСНОСТИ
КАФЕДРА ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ И СИСТЕМ**

ПРОЕКТИРОВАНИЕ ПОИСКОВЫХ МАШИН В ИНТЕРНЕТ

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

По направлению подготовки 09.04.03 Прикладная информатика
направленность: Управление данными и знаниями в компьютерных сетях

Уровень высшего образования: магистратура

Форма обучения: очная, заочная

РПД адаптирована для лиц
с ограниченными возможностями
здоровья и инвалидов

Москва 2024

ПРОЕКТИРОВАНИЕ ПОИСКОВЫХ МАШИН В ИНТЕРНЕТ

Рабочая программа дисциплины

Составитель:

профессор кафедры ИТиС д.т.н. Белый А.Ф.

старший преподаватель кафедры Охупкина Е.П.

Ответственный редактор

к.с.-х.н., доцент,

зав. кафедрой информационных технологий и систем

Н.Ш. Шукенбаева

УТВЕРЖДЕНО

Протокол заседания

кафедры информационных технологий и систем РГГУ

№ 8 от 04.03.2024 г.

ОГЛАВЛЕНИЕ

1	Пояснительная записка.....	4
1.1	Цель и задачи дисциплины.....	4
1.2	Перечень планируемых результатов обучения по дисциплине, соотнесенных с индикаторами достижения компетенций:.....	4
1.3	Место дисциплины в структуре образовательной программы.....	5
2.	Структура дисциплины.....	6
3.	Содержание дисциплины.....	7
4.	Образовательные технологии.....	9
5.	Оценка планируемых результатов обучения.....	9
5.1.	Система оценивания.....	9
5.3.	Оценочные средства (материалы) для текущего контроля успеваемости, промежуточной аттестации обучающихся по дисциплине.....	11
6.	Учебно-методическое и информационное обеспечение дисциплины.....	13
6.1.	Список источников и литературы.....	13
6.2.	Перечень ресурсов информационно-телекоммуникационной сети «Интернет».....	14
6.3	Профессиональные базы данных и информационно-справочные системы.....	14
7.	Материально-техническое обеспечение дисциплины.....	14
8.	Обеспечение образовательного процесса для лиц с ограниченными возможностями здоровья.....	14
9.	Методические материалы.....	17
9.1.	Планы практических занятий и методические указания по их организации и проведению.....	17
	Приложение 1. Аннотация дисциплины.....	34

1 Пояснительная записка

1.1 Цель и задачи дисциплины

Целью освоения дисциплины «Проектирование поисковых машин в Интернет» является теоретическое и практическое освоение методов и технологий создания, формирования и ведения современных баз данных с использованием информационно - поисковых систем, их технологий и языковых средств, а также особенностей поиска информации с их помощью и аналогичными поисковыми возможностями поисковых машин Интернет, с учётом особенностей поиска информации с их помощью и применение этих навыков в различных областях деятельности.

Задачи дисциплины: раскрыть эволюцию исследований в области информационного поиска; охарактеризовать понятие информационного поиска и классификацию его видов; рассмотреть методические основы информационного поиска; охарактеризовать особенности и средства поиска информации в глобальных сетях.

1.2 Перечень планируемых результатов обучения по дисциплине, соотнесенных с индикаторами достижения компетенций:

Коды компетенции	Содержание компетенций	Перечень планируемых результатов обучения по дисциплине
ПК-3 Способен проектировать информационные процессы и системы с использованием инновационных инструментальных средств	ПК-3.1 Знает инновационные инструментальные средства ИТ-сферы	Знать: особенности информационно – поисковых систем (ИПС), как мощного программного средства создания, ведения баз данных и организации поиска в них данных; технологии сбора, накопления, обработки, передачи и распространения информации; технологии, применяемые в ИПС для создания, ведения баз данных и обслуживания абонентов на их основе; языковые средства поиска информации с использованием ИПС и поисковых машин, различия между СУБД, ИПС и поисковыми машинами.
	ПК-3.2 Умеет проектировать информационные процессы и системы с использованием инновационных инструментальных средств	Уметь: организовывать ведение словарей в ИПС, применять языковые средства ИПС и поисковых машин для поиска информации в информационных массивах; проектировать и создавать базы данных и управляющие ими приложения на основе ИПС, разрабатывать технологическую документацию; организовывать ведение словарей в ИПС; оценивать качество и затраты проекта
	ПК-3.3 Владеет навыками проектирования информационных процессов и систем с	Владеть: навыками работы с инструментальными средствами ИПС; навыками формирования формы экранов для обслуживания пользователей на основе баз данных ИПС; средствами поиска

	использованием инновационных инструментальных средств	информации в диалоговом и пакетном режимах, в том числе с использованием постоянно – действующих запросов.
ПК-6 Способен формировать стратегию информатизации прикладных процессов и создания прикладных информационных систем в соответствии со стратегией развития предприятий	ПК-6.1 Знает теоретические основы стратегического управления предприятием и информационными технологиями	Знать: знать основные современные программные изделия класса ИПС и основные поисковые машины; методы информационного обслуживания; назначение и виды ИКТ.
	ПК-6.2 Умеет анализировать потребности предприятия в информатизации, планировать развитие ИТ по направлениям	Уметь: использовать необходимые функциональные и технологические стандарты; использовать документальные информационные системы и поисковые машины для информационного обслуживания, а также для проектирования ИС и архивов; применять языковые средства ИПС и поисковых машин для поиска информации в информационных массивах, создавать пользовательскую документацию.
	ПК-6.3 Владеет навыками формирования стратегии информатизации предприятия в соответствии со стратегией развития	Владеть: навыками использования необходимых функциональных и технологических стандартов; навыками разработки технологической документации; навыками индексирования информационных массивов на ПЭВМ с использованием поисковых машин.

1.3. Место дисциплины в структуре образовательной программы.

Дисциплина «Проектирование поисковых машин в Интернет» является элективной дисциплиной части, формируемой участниками образовательных отношений, блока 1 учебного плана по направлению подготовки 09.04.03 Прикладная информатика, направленность: Управление данными и знаниями в компьютерных сетях. Дисциплина реализуется на факультете информационных систем и безопасности кафедрой информационных технологий и систем. Для освоения дисциплины необходимы знания, умения и владения, сформированные в ходе изучения следующих дисциплин: Архитектура предприятий и информационных систем, Современные системы управления базами данных.

В результате освоения дисциплины формируются знания, умения и владения, необходимые для прохождения преддипломной практики, выполнения и защиты выпускной квалификационной работы.

2. Структура дисциплины

Общая трудоёмкость дисциплины составляет 3 з.е., 108 академических.

Структура дисциплины для очной формы обучения

Объем дисциплины в форме контактной работы обучающихся с педагогическими работниками и (или) лицами, привлекаемыми к реализации образовательной программы на иных условиях, при проведении учебных занятий:

Семестр	Тип учебных занятий	Количество часов
4	Лекции	14
4	Практические работы	16
Всего:		30

Объем дисциплины (модуля) в форме самостоятельной работы обучающихся составляет 78 академических часов.

Структура дисциплины для заочной формы обучения

Объем дисциплины в форме контактной работы обучающихся с педагогическими работниками и (или) лицами, привлекаемыми к реализации образовательной программы на иных условиях, при проведении учебных занятий:

Семестр	Тип учебных занятий	Количество часов
4	Лекции	4
5		2
5	Практические занятия	8
Всего:		12

Объем дисциплины (модуля) в форме самостоятельной работы обучающихся составляет 96 академических часов.

3. Содержание дисциплины

№	Наименование раздела дисциплины	Содержание
1.	Основные положения построения и функционирования информационно – поисковых систем.	<p>Назначение и функции ИПС. Основные отличия ИПС от СУБД. Пакет прикладных программ CDS ISIS. Компания DIALOG и ИПС STAIRS. Модели данных, с которыми оперируют в ИПС. Сеанс работы в ИПС и организация поиска за сеанс. Вывод результатов поиска. Первые ИПС в России. ППП ДИАЛОГ – результат доработки и развития ИПС STAIRS. ППП ИРБИС – развитие CDS ISIS. Первые отечественные изделия – Лингвист – 2, ИПС Стрелка. Создание ИПС с использованием нечётких множеств – ИПС СОВА. Информационно – поисковая система Диана, базирующаяся на составлении и использовании для поиска словарей нормализованных словосочетаний. Возникновение поисковых машины Интернет YANDEX, GOOGLE и др., их отличие от ИПС. Появление на основе протокола Z39.50 технологий интеграции поиска документов в распределённых информационных ресурсах сети Интернет.</p>
2.	Функционально – технологические особенности ИПС. (на примере ИПС ДИАЛОГ).	<p>Происхождение документальных АИС в России. Разработки МЦНТИ. Назначение и основные функции ИПС ДИАЛОГ. Навигация при ручном поиске документальной информации. Классические технологии функционирования ИПС. Ретроспективный поиск. Избирательное распределение информации. Логическая структура документальной БД.</p> <p>Поисковые возможности ИПС ДИАЛОГ, язык запросов - стандарт ИСО 8777 всех ИПС. Текстовый поиск. Форматный поиск. Использование тезауэра. Просмотр найденных документов. Физическая структура и навигация в документальной БД. Обработка входного потока документов (загрузка БД). Описание форматных и текстовых полей.</p> <p>Организация диалогового поиска в терминальной сети. Функционально – технологическая схема использования ППП СИС для создания терминальных сетей и сетей клиент – сервер с целью обеспечения доступа к БД ИПС. Миграция этих технологий IBM на ПЭВМ.</p>
3.	Принципы и приемы информационного поиска в WEB – пространстве.	<p>Основные функции информационно-поисковых систем и модели поиска информации. Булева модель поиска. Классическая булева модель. Нормальная конъюнктивно – дизъюнктивная и дизъюнктивно - конъюнктивная формы запросов. Архитектура поисковых систем, соответствующих булевой модели (Диалог, CDS ISIS). Алгоритм поиска. Расширенная булева модель. Модель нечеткого поиска. Векторно-пространственная модель.</p> <p>Вероятностная модель. Характеристики информационного поиска. Полнота и точность, пертинентность. Вычисление показателей качества поиска. Технологические характеристики поисковых систем.</p>
4.	Современные ИПС, применяемые для создания	<p>Интегрированная система "Ирбис", Описание системы, АРМ абонента, организация поиска, просмотр документов. Логическая структура данных ИРБИС. Физическая структура</p>

	информационных систем.	<p>базы данных ИРБИС. Информационно-поисковый язык документальной ИПС ИРБИС. Синтаксис и семантика использования дескрипторов. Средства формирования запросов. Вытеснение произвольных внутрисистемных форматов MARC–ориентированными форматами.</p> <p>ИПС РС BIRS, её возможности, использование на ПЭВМ, возможность ведения больших полнотекстовых массивов и организация поиска в этой ИПС.</p>
5.	Проектирование поисковых машины Интернет	<p>Виды данных Интернет. Классификация информационных ресурсов Интернет. Классификация Web ресурсов: навигационные сайты и конечные сайты. Организация поиска информации в мировой сети. Поисковые машины и каталоги. Приемы и методы проектирования. Организация хранения данных: данные в реляционных хранилищах, многомерное хранение данных, хранилища данных. Компоненты хранилищ данных. Визуализация многомерных данных. Модели данных для индексации текстовых и графических информационных ресурсов. Информационные агентства и каталоги. Поисковые машины, как поисковые системы Интернет их отличия от ИПС. Поисковые машины Апорт, Яндекс, Рамблер и Google, особенности их языковых средств. Глобальные информационные базы данных. Модель WEB – пространства. Статическая и динамическая составляющие. Недостатки поисковых средств поисковых машин. Развитие поисковых возможностей ИПС Интернет. Концепция Text Mining. Концепция WIKI и её реализация. Открытые энциклопедии Интернет. Стратегии поиска информации в Интернет.</p>
6.	Технология функционирования ИПС в распределенной информационной среде сети Интернет	<p>Суть стандарта Z39.50. Z ассоциация – сеанс ИПС. Основное отличие от поисковых машин и технологии поиска. Службы, обеспечивающие реализацию сервисов протокола Z39.50. Структура программного обеспечения. Основные производители серверных программных изделий. Отечественные библиотечные сети Z39.50. ИПС «Руслан».</p>

4. Образовательные технологии

Для проведения учебных занятий по дисциплине используются различные образовательные технологии. Для организации учебного процесса может быть использовано электронное обучение и (или) дистанционные образовательные технологии.

5. Оценка планируемых результатов обучения

5.1. Система оценивания

Форма контроля	Макс. количество баллов	
	За одну работу	Всего
Текущий контроль:		
Практическая работа № 1, защита отчета	6 баллов	60 баллов
Практическая работа № 2, защита отчета	9 баллов	
Практическая работа № 3, защита отчета	9 баллов	
Практическая работа № 4, защита отчета	9 баллов	
Практическая работа № 5, защита отчета	9 баллов	
Практическая работа № 6, защита отчета	9 баллов	
Практическая работа № 7, защита отчета	9 баллов	
Промежуточная аттестация <i>зачет с оценкой</i>		40 баллов
Итого за семестр		100 баллов

Полученный совокупный результат конвертируется в традиционную шкалу оценок и в шкалу оценок Европейской системы переноса и накопления кредитов (European Credit Transfer System; далее – ECTS) в соответствии с таблицей:

100-балльная шкала	Традиционная шкала		Шкала ECTS
95 – 100	Отлично	зачтено	A
83 – 94			B
68 – 82	Хорошо		C
56 – 67	Удовлетворительно		D
50 – 55			E
20 – 49	Неудовлетворительно		FX
0 – 19			не зачтено

5.2. Критерии выставления оценки по дисциплине

Баллы/ Шкала ECTS	Оценка по дисциплине	Критерии оценки результатов обучения по дисциплине
100-83/ A,B	«отлично»	Выставляется обучающемуся, если он глубоко и прочно усвоил теоретический и практический материал, может продемонстрировать это на занятиях и в ходе промежуточной аттестации. Обучающийся исчерпывающе и логически стройно излагает учебный материал, умеет увязывать теорию с практикой, справляется с решением задач профессиональной направленности высокого уровня сложности, правильно обосновывает принятые решения.

Баллы/ Шкала ECTS	Оценка по дисциплине	Критерии оценки результатов обучения по дисциплине
		<p>Свободно ориентируется в учебной и профессиональной литературе.</p> <p>Оценка по дисциплине выставляются обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции, закреплённые за дисциплиной, сформированы на уровне – «высокий».</p>
82-68/ С	«хорошо»	<p>Выставляется обучающемуся, если он знает теоретический и практический материал, грамотно и по существу излагает его на занятиях и в ходе промежуточной аттестации, не допуская существенных неточностей.</p> <p>Обучающийся правильно применяет теоретические положения при решении практических задач профессиональной направленности разного уровня сложности, владеет необходимыми для этого навыками и приёмами.</p> <p>Достаточно хорошо ориентируется в учебной и профессиональной литературе.</p> <p>Оценка по дисциплине выставляются обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции, закреплённые за дисциплиной, сформированы на уровне – «хороший».</p>
67-50/ D,E	«удовлетворительно»	<p>Выставляется обучающемуся, если он знает на базовом уровне теоретический и практический материал, допускает отдельные ошибки при его изложении на занятиях и в ходе промежуточной аттестации.</p> <p>Обучающийся испытывает определённые затруднения в применении теоретических положений при решении практических задач профессиональной направленности стандартного уровня сложности, владеет необходимыми для этого базовыми навыками и приёмами.</p> <p>Демонстрирует достаточный уровень знания учебной литературы по дисциплине.</p> <p>Оценка по дисциплине выставляются обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции, закреплённые за дисциплиной, сформированы на уровне – «достаточный».</p>
49-0/ F,FX	«неудовлетворительно»	<p>Выставляется обучающемуся, если он не знает на базовом уровне теоретический и практический материал, допускает грубые ошибки при его изложении на занятиях и в ходе промежуточной аттестации.</p> <p>Обучающийся испытывает серьёзные затруднения в применении теоретических положений при решении практических задач профессиональной направленности</p>

Баллы/ Шкала ECTS	Оценка по дисциплине	Критерии оценки результатов обучения по дисциплине
		<p>стандартного уровня сложности, не владеет необходимыми для этого навыками и приёмами. Демонстрирует фрагментарные знания учебной литературы по дисциплине.</p> <p>Оценка по дисциплине выставляются обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции на уровне «достаточный», закреплённые за дисциплиной, не сформированы.</p>

5.3. Оценочные средства (материалы) для текущего контроля успеваемости, промежуточной аттестации обучающихся по дисциплине

Контрольные вопросы для зачета (ПК-3, ПК-6)

1. Перечислите основные способы организации хранения и поиска документов.
2. Что такое информационно-поисковая система (ИПС)?
3. Особенности предметной области документальных систем.
4. Основные структуры данных ИПС, её архитектура.
5. Принципы и приемы проектирования ИПС.
6. Сопоставление ИПС и СУБД.
7. Поисковые системы Интернет на базе протокола Z39.50.
8. Сравнение технологии работы абонента Интернет по протоколу Z39.50 и HTTP.
9. Назначение и основные функции ИПС ДИАЛОГ.
10. Ретроспективный поиск ИПС ДИАЛОГ.
11. Избирательное распределение информации.
12. Логическая структура документальной БД.
13. Язык запросов ИПС ДИАЛОГ.
14. Назначение и использование тезауруса в ИПС ДИАЛОГ.
15. Физическая структура и навигация в документальной БД.
16. Обработка входного потока документов (загрузка БД)
17. Архитектура «клиент/сервер» и «терминальная сеть», их реализация для доступа к БД ИПС.
18. Основные функции информационно-поисковых систем и модели поиска информации.
19. Булева модель поиска.
20. Нормальная конъюнктивно – дизъюнктивная и дизъюнктивно - конъюнктивная формы запросов.
21. Архитектура поисковых систем, соответствующих булевой модели (Диалог, CDS ISIS).
22. Алгоритм поиска.
23. Что такое избыточное индексирование? Какими преимуществами и недостатками оно обладает?
24. Основные положения модели нечеткого поиска.
25. Характеристики информационного поиска - полнота и точность, пертинентность.

Тематика рефератов (докладов) (ПК-3, ПК-6)

1. Технологические характеристики поисковых систем.
2. Интегрированная система ИРБИС, её основные свойства и архитектура.
3. Логическая структура данных ИРБИС.

4. Физическая структура базы данных ИРБИС.
5. Особенности внутрисистемного формата MARC.
6. Функциональные особенности ИПС РС BIRS
7. Перечислите основные функции (стандартные возможности) информационно-библиотечных систем.
8. Корпоративные автоматизированные библиотечно-информационные технологии: общая характеристика.
9. Отличия поисковых машин и ИПС.
10. Поисковые машины, как поисковые системы Интернет их технологические особенности.
11. Поисковые машины Апорт, Яндекс, Рамблер и Google, особенности их языковых средств.
12. Глобальные информационные базы данных.
13. В чем заключена специфика использования ИПС в Интернете?
14. Витринная функция Интернета в деятельности библиотек
15. Топология Web-пространства
16. WIKI - корпоративная технология и Википедия.
17. Разметка документов в WIKI.
18. Text Mining общее представление.
19. Суть стандарта Z39.50.
20. Сеанс ИПС, использующей стандарт Z39.50, его основное отличие от поисковых машин и технологии поиска.
21. Службы, обеспечивающие реализацию сервисов протокола Z39.50.
22. Структура программного обеспечения.
23. Отечественные библиотечные сети Z39.50.

6. Учебно-методическое и информационное обеспечение дисциплины

6.1. Список источников и литературы

Литература

Основная

1. Гаврилова, Т. А. Инженерия знаний. Модели и методы / Т. А. Гаврилова, Д. В. Кудрявцев, Д. И. Муромцев. — 6-е изд., стер. — Санкт-Петербург : Лань, 2023. — 324 с. — ISBN 978-5-507-46580-4. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/312842> (дата обращения: 31.03.2024). — Режим доступа: для авториз. пользователей.
2. Волкова, В. Н. Теория информационных процессов и систем : учебник и практикум для вузов / В. Н. Волкова. — 2-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2023. — 432 с. — (Высшее образование). — ISBN 978-5-534-05621-1. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/511112>.
3. Гасанов, Э. Э. Интеллектуальные системы. Теория хранения и поиска информации : учебник для вузов / Э. Э. Гасанов, В. Б. Кудрявцев. — 2-е изд., испр. и доп. — Москва : Издательство Юрайт, 2023. — 271 с. — (Высшее образование). — ISBN 978-5-534-08684-3. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/513151>.
4. Станкевич, Л. А. Интеллектуальные системы и технологии : учебник и практикум для вузов / Л. А. Станкевич. — 2-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2023. — 495 с. — (Высшее образование). — ISBN 978-5-534-16238-7. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/530657>.
5. 2. Землянский, А. А. Управление информационными ресурсами в научно-исследовательской работе: учебное пособие / А. А. Землянский, И. Е. Быстренина. - 2-е изд. - Москва: Дашков и К, 2021. - 110 с. - URL: <https://znanium.com/catalog/product/1232484>.)

Дополнительная

6. Мартишин, С. А. Базы данных. Практическое применение СУБД SQL и NoSQL-типа для проектирования информационных систем: учеб. пособие / С.А. Мартишин, В.Л. Симонов, М.В. Храпченко. — Москва: ИД «ФОРУМ»: ИНФРА-М, 2019. — 368 с. - URL: <https://znanium.com/catalog/product/1001370>.
7. Шитиков Владимир Кириллович. Классификация, регрессия и другие алгоритмы Data Mining с использованием R [Электронный ресурс] / В. К. Шитиков, С. Э. Мастицкий. - Электрон. дан. - Тольятти; Лондон : [б. и.], 2017. - 351 с. Ссылка на ресурс: <https://github.com/ranalytics/data-mining>
8. Мартыненко, Г. Я. Основы стилеметрии: учебно-методическое пособие / Г. Я. Мартыненко, А. О. Гребенников. - Санкт-Петербург: СПбГУ, 2018. - 27 с. - URL: <https://znanium.com/catalog/product/1243852>.
9. Перебейнос Валентина Исидоровна. Статистические методы для лингвистов / В. И. Перебейнос; [под ред. М. Р. Кауль ; Рос. гос. гуманитарный ун-т]. - Москва: РГГУ, 2016. — 238 с. (24 экз.).
10. Волосатова, Т. М. Информатика и лингвистика: учебное пособие / Т.М. Волосатова, Н.В. Чичварин. — Москва: ИНФРА-М, 2021. — 196 с. - URL: <https://znanium.com/catalog/product/1212524>.
11. Казарин, Ю. В. Лингвистический анализ текста: учебное пособие для вузов / Ю.В. Казарин; под научной редакцией Л.Г. Бабенко. — 2-е изд. — Москва: Издательство Юрайт, 2020. — 132 с.— URL: <https://urait.ru/bcode/454651>.

12. Функциональная грамматика русского языка: учебник и практикум для вузов / С. М. Колесникова [и др.]; под редакцией С. М. Колесниковой. — Москва: Издательство Юрайт, 2020. — 421 с. — (Высшее образование). — ISBN 978-5-534-12882-6. — Текст: электронный // ЭБС Юрайт [сайт]. — URL: <https://urait.ru/bcode/460524>

6.2. Перечень ресурсов информационно-телекоммуникационной сети «Интернет».

1. Об индексировании поисковыми машинами сайтов на фреймах // <https://www.ashmanov.com/> URL: <https://www.ashmanov.com/education/articles/2-ob-indeksirovanii-poiskovymi-mashinami-saytov-na-freymakh/>.
2. Построение индекса для поисковой машины // habr.com URL: <https://habr.com/ru/post/124728/>.
3. Создание собственной поисковой системы с помощью PHP // [ibm.com](https://www.ibm.com) URL: <https://www.ibm.com/developerworks/ru/library/os-php-sphinxsearch/>.

12.3 Профессиональные базы данных и информационно-справочные системы

Доступ к профессиональным базам данных: <https://liber.rsuh.ru/ru/bases>

Информационные справочные системы:

1. Консультант Плюс
2. Гарант

7. Материально-техническое обеспечение дисциплины

№п/п	Наименование специальных помещений и помещений для самостоятельной работы	Оснащенность специальных помещений и помещений для самостоятельной работы	Перечень лицензионного программного обеспечения. Реквизиты подтверждающего документа		
			Наименование ПО	Лицензия/сертификат/заказ	Дата лицензии
1.	Лаборатория информатики – ауд. № 203	1 компьютер преподавателя, 12 компьютеров обучающихся, маркерная доска, проектор	Windows 7 Microsoft office 2010 Pro Microsoft Visual Professional 2019 Mozilla Firefox 52.8.1 ESR Matlab Mathcad Education - University edition Kaspersky Endpoint Security	68526624 49420326 63202190 свободный доступ 647526 2996385 17E0-181226-094912-873-979	без даты 08.12.2011 без даты свободный доступ без даты 14.06.2019 26.12.2018

8. Обеспечение образовательного процесса для лиц с ограниченными возможностями здоровья

В ходе реализации дисциплины используются следующие дополнительные

методы обучения, текущего контроля успеваемости и промежуточной аттестации обучающихся в зависимости от их индивидуальных особенностей:

- для слепых и слабовидящих:
 - лекции оформляются в виде электронного документа, доступного с помощью компьютера со специализированным программным обеспечением;
 - письменные задания выполняются на компьютере со специализированным программным обеспечением, или могут быть заменены устным ответом;
 - обеспечивается индивидуальное равномерное освещение не менее 300 люкс;
 - для выполнения задания при необходимости предоставляется увеличивающее устройство; возможно также использование собственных увеличивающих устройств;
 - письменные задания оформляются увеличенным шрифтом;
 - экзамен и зачёт проводятся в устной форме или выполняются в письменной форме на компьютере.
- для глухих и слабослышащих:
 - лекции оформляются в виде электронного документа, либо предоставляется звукоусиливающая аппаратура индивидуального пользования;
 - письменные задания выполняются на компьютере в письменной форме;
 - экзамен и зачёт проводятся в письменной форме на компьютере; возможно проведение в форме тестирования.
- для лиц с нарушениями опорно-двигательного аппарата:
 - лекции оформляются в виде электронного документа, доступного с помощью компьютера со специализированным программным обеспечением;
 - письменные задания выполняются на компьютере со специализированным программным обеспечением;
 - экзамен и зачёт проводятся в устной форме или выполняются в письменной форме на компьютере.

При необходимости предусматривается увеличение времени для подготовки ответа.

Процедура проведения промежуточной аттестации для обучающихся устанавливается с учётом их индивидуальных психофизических особенностей. Промежуточная аттестация может проводиться в несколько этапов.

При проведении процедуры оценивания результатов обучения предусматривается использование технических средств, необходимых в связи с индивидуальными особенностями обучающихся. Эти средства могут быть предоставлены университетом, или могут использоваться собственные технические средства.

Проведение процедуры оценивания результатов обучения допускается с использованием дистанционных образовательных технологий.

Обеспечивается доступ к информационным и библиографическим ресурсам в сети Интернет для каждого обучающегося в формах, адаптированных к ограничениям их здоровья и восприятия информации:

- для слепых и слабовидящих:
 - в печатной форме увеличенным шрифтом;
 - в форме электронного документа;
 - в форме аудиофайла.
- для глухих и слабослышащих:
 - в печатной форме;
 - в форме электронного документа.
- для обучающихся с нарушениями опорно-двигательного аппарата:
 - в печатной форме;
 - в форме электронного документа;
 - в форме аудиофайла.

Учебные аудитории для всех видов контактной и самостоятельной работы, научная библиотека и иные помещения для обучения оснащены специальным оборудованием и учебными местами с техническими средствами обучения:

- для слепых и слабовидящих:
 - устройством для сканирования и чтения с камерой SARA CE;
 - дисплеем Брайля PAC Mate 20;
 - принтером Брайля EmBraille ViewPlus;
- для глухих и слабослышащих:
 - автоматизированным рабочим местом для людей с нарушением слуха и слабослышащих;
 - акустический усилитель и колонки;
- для обучающихся с нарушениями опорно-двигательного аппарата:
 - передвижными, регулируемыми эргономическими партами СИ-1;
 - компьютерной техникой со специальным программным обеспечением.

9. Методические материалы

9.1. Планы практических занятий и методические указания по их организации и проведению

Практическая работа 1. Лингвистические средства представления смыслового содержания

1. Цель работы

Целью работы является изучение лингвистических средств представления смыслового содержания сообщения.

Работа имеет практический характер и способствует пониманию роли лингвистического обеспечения в контексте эффективности поиска информации.

Варианты заданий представлены в форме научных статей в заданной предметной области, содержание которых надо описать рубрикатора, УДК и языком дескрипторного типа.

2. Теоретические сведения

Лингвистическое обеспечение – это совокупность языковых средств (в том числе и правила реферирования и индексирования), позволяющих более или менее подготовленному пользователю взаимодействовать с машинной системой (по крайней мере, в части отыскания нужных пользователю документов).

ЛО имеет еще и «системное», предназначение – согласование точек зрения на предметную область как для различных поставщиков и потребителей информации, так и во времени: ИПЯ имеет искусственную природу и создается в конкретных условиях.

В основу методов представления смысла положена та или иная знаковая система, но при этом различают классификационный и описательный подходы.

Классификация, как средство описания содержания документа, представляет собой процесс соотнесения содержания документов с понятиями, зафиксированными в заранее составленных систематических схемах. Классификационные методы обеспечивают систематизацию объектов в соответствии с некоторой заданной классификационной схемой. Код, присвоенный отдельному классу, обеспечивает его полную идентификацию в рамках конкретного классификатора.

Описательные методы идентификации используются, как правило, в тех случаях, когда необходимо идентифицировать конкретный объект или группу объектов путем описания произвольного набора его характеристик. Описательный метод предполагает наряду с указанием классификационных характеристик выделение дополнительных наборов свойств, углубляющих характеристику объекта и сужающих область поиска. Описательный подход в большей степени поддерживается языками дескрипторного типа. Они реализуют *координатное индексирование*, которое заключается в формировании описания документа как совокупности дескрипторов, выбираемых из заранее созданных словарей понятий либо из текстов документов.

Государственный рубрикатор научно-технической информации

Государственный рубрикатор научно-технической информации (ГРНТИ) представляет собой универсальную трехуровневую иерархическую классификацию областей знания, принятую для систематизации всего потока научно-технической информации.

Рубрикатор имеет многоцелевое назначение, отвечает потребностям всех информационных органов в силу универсальности охвата тематики, обеспечивает многофункциональное использование.

Рубрикам (в качестве справочной информации) поставлены в соответствие индексы УДК и коды Номенклатуры специальностей научных работников ВАК. Это

обеспечивает взаимосвязь между классификационными системами, а также возможность поиска документов в информационных массивах и базах данных, систематизированных по УДК и номенклатуре специальностей ВАК.

Состав и структура Рубрикатора. Рубрикатор имеет 3 уровня иерархии, при этом весь универсум знаний условно разделен на 4 подкласса:

- «Общественные науки» (значение кода первого уровня от 00 до 26);
- «Естественные и точные науки» (значение кода первого уровня от 27 до 43);
- «Технические и прикладные науки. Отрасли экономики» (значение кода первого уровня от 44 до 81);
- «Межотраслевые и комплексные проблемы» (значение кода первого уровня от 82 до 90).

Каждая рубрика состоит из кода (нотации) и наименования (описание класса), а также может иметь при себе ссылки и примечания.

На каждом уровне Рубрикатора возможно деление на 100 подклассов. Коды рубрик состоят из цепочки пар арабских цифр, разделенных точкой. В конце кода точка не ставится.

Наименование рубрики представляет собой текст, отражающий ее смысловое содержание, например:

37.23.31 Моделирование климата. Прогноз климата.

Рубрикатор является иерархической классификацией и построен так, что классы одного уровня, как правило, не пересекаются, а понятия одного уровня находятся в состоянии подчинения к понятию более высокого уровня, например:

73 Транспорт

73.29 Железнодорожный транспорт

73.31 Автомобильный транспорт

73.34 Водный транспорт

73.37 Воздушный транспорт

73.39 Трубопроводный транспорт

Признаком, по которому проведено деление класса «**73** Транспорт», является признак отнесения к видам технических средств перевозки. Перечисленные виды транспорта не пересекаются, а в совокупности составляют класс "Транспорт" — обобщенное понятие более высокого уровня иерархии.

В Рубрикаторе использовано сочетание иерархии с фасетным принципом, который проявляется в наличии совокупности рубрик, повторяющейся в разных классах в виде группы "Общие вопросы", а также в применении типовых классификационных делений в разных разделах Рубрикатора.

Для исключения дублирования разделы Рубрикатора включают тематическое «ядро» предметной области, оснащенное ссылками на связанные с ним смежные вопросы, помещенные в других тематически близких разделах Рубрикатора.

Уровень рубрики, соответствующей определенной области знания, отражает не ее значимость, а только степень обобщения при логической группировке понятий. В пределах одного уровня располагаются примерно равнозначные по объему понятия.

Рубрикатор обеспечивает отнесение каждого документа к минимально необходимому числу рубрик, поскольку между рубриками, имеющими логико-семантические связи, существуют разграничения аспектов, отраженных в наименованиях рубрик, примечаниях и ссылках.

Универсальная десятичная классификация

УДК, построенная по систематическому принципу, понимается как *система цифрового кодирования* не содержания, а *поисковых образов (признаков) документа*, которые формально отделяются от его содержания.

С 1963 г. классификация всех публикаций в области технических и естественных наук по УДК введена в качестве обязательной в научно-технических издательствах, редакциях научно-технических журналов, органах научно-технической информации.

Структура УДК. Все классы УДК сгруппированы в шесть фасетов (Общий предмет, Место, Народность, Время, Язык документа, Форма документа), каждый из которых подразделяется по иерархическому принципу на несколько уровней. Для идентификации отдельного класса на каждом уровне используются десятичные цифры, тем самым, в индексах УДК каждая последующая цифра не меняет значения предыдущих, а лишь уточняет их, обозначая частное понятие. Индекс УДК представляет собой последовательность десятичных цифр, возможно разделенных на группы знаками «точка».

Фасет «Общий предмет» (основная таблица классификации) имеет десять основных подразделений, которые называются *главными классами*.

Основная таблица включает понятия и соответствующие им индексы, охватывающие классы от 0 до 9 с их дальнейшими делениями по содержанию. Главными классами фасета «Общий предмет» являются:

- 0 Общий отдел
- 1 Философия. Психология. Логика
- 3 Общественные науки
- 5 Математика. Естественные науки
- 6 Прикладные науки. Медицина. Техника
- 7 Искусство. Декоративно-прикладное искусство. Фотография. Музыка.

Игры. Спорт

- 8 Языкознание. Филология. Художественная литература.

Литературоведение

- 9 География. Биографии. История

Остальные фасеты УДК получили название *вспомогательных таблиц*, предназначенных для классификации по дополнительным признакам. Вспомогательные индексы (определители) этих таблиц бывают двух видов: *общие*, которые могут соединяться с любыми индексами основной таблицы УДК, и *специальные*, присоединяемые только к понятиям текущего раздела.

Кроме общей фасетной формулы в УДК применяется также субформула для построения сложных предметных индексов. В этом случае предметный индекс УДК состоит из нескольких простых индексов, соединяемых друг с другом знаками присоединения, распространения, объединения или отношения.

Международная патентная классификация

Международная патентная классификация (МПК) обеспечивает достаточно полное индексирование предмета патентования с помощью ограниченного числа рубрик за счет ориентации последних на аспекты, важные с точки зрения патентного права, такие как характеристика вещества или устройства, его функции, применение, получение. Классификация представляет собой специальную, имеющую технико-прикладной характер, линейную систему иерархического типа, предназначенную для ручного индексирования. Согласно основным принципам применения Международной патентной классификации, указывающим, что изобретение, подлежащее классификации, не может

рассматриваться как чистая идея в отрыве от ее технического воплощения в устройстве, способе или веществе, выделяют следующие предметы (объекты) изобретения:

- 1) вещество или материал;
- 2) устройство, прибор, конструкция;
- 3) процесс, способ, метод;

Правила классифицирования. Основным назначением МПК является облегчение поиска аналогичных технических решений. В связи с этим МПК разработана и должна использоваться таким образом, чтобы одинаковые технические решения классифицировались одинаково, и, следовательно, могли быть найдены в одном и том же месте системы классификации. *Методика классифицирования* опирается на выявление технической сущности изобретения. Техническую сущность изобретения может представлять собой способ, продукт или устройство (или способ их использования или применения), причем эти термины используются в МПК в самом широком смысле. Например:

способ – полимеризация, ферментация, разделение, формование, транспортирование, обработка текстиля, передача и преобразование энергии, строительство, методы работы на машинах, способы работы машин, обработка и передача информации;

продукт - химическое соединение, состав, ткань;

устройство - установка для проведения химических или физических процессов, инструмент, орудие, машина, устройство для выполнения технологических операций.

Структура МПК. МПК охватывает все области знаний, которые могут подлежать защите охраняемыми документами и имеет иерархическую 4-х уровневую структуру.

1. *Раздел.* Обозначается латинской заглавной буквой от А до Н и снабжен *заголовком*, укрупнено отражающим его содержание.

МПК включает следующие 8 разделов:

А - удовлетворение жизненных потребностей человека;

В - различные технологические процессы, транспортирование;

С - химия, металлургия;

Д - текстиль, бумага;

Е - строительство, горное дело;

Ф - механика, освещение, отопление, двигатели и насосы, оружие, боеприпасы, взрывные работы;

Г - физика;

Н - электричество.

2. *Класс.* Каждый раздел делится на классы, обозначаемые двузначным числом. Содержание класса отражает *заголовок класса*. Некоторые классы снабжаются кратким перечнем относящейся к ним тематики - *указателем класса*.

3. *Подкласс.* Каждый класс содержит один или более подклассов, обозначаемых заглавной буквой латинского алфавита. Содержание подкласса определяет *заголовок подкласса*. Некоторые подклассы снабжаются кратким перечнем относящейся к ним тематики - *указателем содержания подкласса*.

4. *Группа, подгруппа.* Каждый подкласс разбит на подразделения, которые в дальнейшем именуется «дробными рубриками». Среди дробных рубрик различают основные группы и подгруппы.

Дробная рубрика обозначается двумя числами, разделенными наклонной чертой. Первое (максимум трехзначное, обычно нечетное) число индексирует *основную группу*, второе (минимум 2 цифры, обычно четное) – *подгруппу*. Для основной группы код подгруппы имеет значение 00. Каждую третью или четвертую цифру после наклонной черты следует понимать как дальнейшее десятичное деление предыдущей цифры. Отсюда

следует, что подгруппа с индексом 5/417 должна стоять после подгруппы 5/41, но перед подгруппой 4/42.

Текст основной группы определяет область, которая считается целесообразной для проведения поиска. Текст и индексы основных групп выделены жирным шрифтом.

Текст подгруппы понимается всегда в пределах объема ее основной группы и определяет тематическую область, в которой считается целесообразным проведение поиска. Перед текстом подгруппы ставится одна или более точек, которые определяют степень ее подчиненности, т.е. указывают на то, что подгруппа является рубрикой, подчиненной ближайшей вышестоящей рубрике, напечатанной с меньшим сдвигом, т.е. имеющей на одну точку меньше.

Анализ содержания текста документа

Сначала анализируется заголовок, затем реферат, оглавление и полный текст документа.

Из заголовка, реферата, оглавления и текста документа выбираются понятия (ключевые слова), наиболее полно отражающие содержание документа.

При анализе содержания документа следует выбирать только полезную информацию. Критерием, которым следует руководствоваться при этом выборе, является то, что полезная информация наиболее информативна и интересна с точки зрения пользователя системы. Документ может быть адекватно определен одним, двумя и, возможно, тремя наиболее важными аспектами, которые обычно являются основанием для опубликования этого документа. В процессе подготовки поискового образа необходимо представить каждый из этих аспектов одним или несколькими понятиями и затем согласовать эти группы для отражения содержания рассматриваемого документа.

При индексировании публикаций предметными рубриками ГРНТИ или УДК необходимо выделить основной предмет исследований публикации в контексте тематики предметной области. Основной задачей является определение первой рубрики (с точки зрения важности для предметной области, например ядерной науки и технологии). Первая рубрика должна отвечать основной теме исследований, результаты которых рассматриваются в данной конкретной публикации. Если в публикации приведена информация, относящаяся к другим тематическим направлениям, правила рубрикации позволяют приписать ей одну или несколько второстепенных предметных рубрик. Теоретически число второстепенных рубрик не ограничено, однако их число обычно не должно превышать одной или двух.

При выборе ключевых слов следует помнить, что следует использовать только те ключевые слова, которые действительно отражают понятия, рассматриваемые в документе, т.е. каждое выбранное ключевое слово должно нести достаточную информативную нагрузку.

3. Порядок выполнения работы

3.1. Провести содержательный анализ документа – варианта индивидуального задания. Определить основной предмет и аспекты.

3.2. Построить систему (дерево) основных понятий.

3.3. Сформировать поисковые образы документа для каждого аспекта с использованием ГРНТИ, УДК и/или МПК, ключевых слов.

Практическая работа 2. Проектирование представления данных документальной ИПС

1. Цель работы

Целью работы является изучение лингвистических средств представления записей в документальной БД.

Работа имеет практический характер и способствует пониманию роли языков определения данных в контексте эффективности представления и использования материала документальной ИПС.

Варианты заданий представлены в форме наборов элементов данных библиографических описаний отдельного вида публикаций.

2. Теоретические сведения

Различные практические цели использования документальной информации (например, составление библиографического списка цитированной литературы, тематический поиск или анализ виртуальных авторских коллективов) определяют необходимость управление формой представления документов и/или списков документов. В документальных АИС, где разнообразие информации достаточно ограничено, для определения формата представления обычно используются так называемые схемы документов (или схемы БД), позволяющие оперативно изменить форму и наполнение отображения документа.

3. Порядок выполнения работы

Для заданной формы библиографического описания вида издания сформировать схему документа, определив элементы описания и команды управления выводом (Приложение 1).

Приложение 1

Описание полей схемы документа АИС IRBIS

Описатель	Значение	Комментарий
имя поля #имя_базы_данных *IDN FILLER, TRAILER	строка до 11 прописных символов	Алфавитно-цифровое значение, начинающееся всегда с буквы, которое используется в формулировках запроса или для выбора словаря для просмотра. Имя поля, равное '#имя_базы_данных', специфицирует подстановку текста документов, найденных в указанной базе данных по запросу-подстановке, имя файла которого приведено в описателе заголовка поля. Текст будет сформатирован по схеме, имя которой приведено в описателе управления выводом. Имя поля, равное '*IDN', специфицирует вывод внутреннего идентификатора документа. Для расширения возможностей отображения документов могут использоваться вспомогательные поля-заполнители: FILLER и TRAILER. Если в качестве имени поля указывается значение FILLER, то будет выводиться только текст заголовка в соответствии с командами управления выводом. Если в имени поля используется значение TRAILER, то обработка аналогична FILLER-полю, но только при условии, что поле документа, непосредственно предшествующее в описании TRAILER-полю, было непустое. Если значение метки поля-заполнителя не

INDEX		равно 0, то вывод заголовка будет производиться в том случае, если поле документа с меткой поля, равной специфицированной, будет не пустое. Если в имени поля используется значение INDEX, то при вызове функции F6 просмотра словаря значение этого поля будет выводиться только для первого документа списка документов, содержащих слово, на которое указывает маркер.
LINKUP, LINKDOWN		Имя поля, равное 'LINKUP' ('LINKDOWN'), для поля ссылочного типа используется для физической связи с предыдущим (последующим) физическим документом, переход к которому происходит при двойном нажатии клавиши перемещения маркера вверх (вниз), если последний уже находится в первой (последней) строке документа. Значения этих полей должны быть сформированы при загрузке документов в базу данных.
&псевдоним		Имя поля равное '&псевдоним' при загрузке схемы будет заменено на фактическое значение имени, заданное в файле конфигурации БД в строке определения псевдонима: &псевдоним = факт.имя поля;
\$FCOMMENT		Имя поля равное \$FCOMMENT определяет Наименование этого поля в списке поисковых полей вместо заголовка.
заголовок поля	строка до 19 символов	Текст, предшествующий при отображении материалу поля. Знак '\' в заголовке вызывает размещение текста, следующего за этим знаком, с новой строки.
метка сегмента	1<=число<=1 27	Десятичное число, идентифицирующее материал поля в базе данных.
флаг загрузки	Y/N	Символьное значение, разрешающее или запрещающее загрузку в базу данных. Материал поля, для которого флаг имеет значение 'Y', при загрузке помещается в базу данных; если же флаг имеет значение 'N', материал поля игнорируется в момент загрузки.
формат	A/N/E/B/H/R/ D/J/I/S/P/T	Символьное значение, задающее способ представления материала. Материал поля может быть представлен: - в символьном формате ('A'); - в цифровом формате ('N'); - в числовом формате с плавающей точкой ('E'); - битовой последовательностью ('B'); - байтовой последовательностью в шестнадцатеричном представлении ('H'); - в виде ссылки на документ ('R'); - в виде даты в формате YYMMDD ('D'); - в виде даты в формате DD month YYYY ('J'); - в виде даты в формате DD.MM.YY ('I'); - в виде даты в формате MM-DD-YY ('S'); - графическом файлом в формате PCX ('P'), или в формате TIFF ('T'). Если поле представлено в цифровом формате ('N'), то соответствующее значение в словаре будет дополнено до указанной длины лидирующими нулями, а при поиске сравнение будет проводиться по правилам сравнения целых чисел.
длина поля	0/число	Число, задающее для поля фиксированной длины количество позиций, резервируемых для размещения

		<p>материала.</p> <p>Для полей неограниченной длины необходимо задать нулевое значение.</p> <p>Для цифровых полей каждое значение в словаре будет дополнено до заданной длины нулями слева.</p>
ключ	N/K/D/A	<p>Символьное значение, задающее характер использования справочных структур.</p> <p>Для полей со значением ключа 'N' справочные поисковые структуры не создаются.</p> <p>Для полей со значениями ключа 'K' и 'D' строятся справочные поисковые структуры, причем имя поля, имеющего значение ключа 'D', при открытии базы данных будет установлено как текущее и, если оно не изменяется другими средствами, будет использоваться в любом запросе, требующем имя поля по умолчанию.</p> <p>Для полей со значением ключа 'A' при поиске используются справочные структуры, построенные по материалу возможно нескольких полей.</p>
метка ключа	1<=число<=1 27	<p>Десятичное число, идентифицирующее при загрузке любое поле, материал которого использовался для построения словаря, а при поиске задающее внутренний идентификатор словаря, по которому будет проводиться отбор документов по критерию, использующему это поле.</p>
строчный регистр	Y/N	<p>Символьное значение, определяющее характер использования строчного регистра в дескрипторах словаря.</p> <p>Значение 'Y' разрешает использование строчных букв. Если задано значение 'N', то строчные буквы в дескрипторах словаря и ключевых словах запроса преобразуются в прописные.</p>
разделители предложений	список до 11 символов	<p>Список символов, рассматриваемых в качестве ограничителей предложений при выделении слов на этапе построения словарей, а также при поиске и отображении материала.</p> <p>Для спецификации разделителей с кодами меньше 32 необходимо использовать двухсимвольную конструкцию, в которой код разделителя получается путем вычитания кода символа '@' (шестнадцатеричное 40) из кода второго символа.</p> <p>Двойной символ '@@' дает значение разделителя '@'.</p>
разделители слов	список до 11 символов	<p>Список символов, рассматриваемых в качестве ограничителей при выделении слов на этапе построения словарей, а также при поиске и отображении материала.</p> <p>Для спецификации разделителей с кодами меньше 32 необходимо использовать двухсимвольную конструкцию, в которой код разделителя получается путем вычитания кода символа '@' (шестнадцатеричное 40) из кода второго символа.</p> <p>Двойной символ '@@' дает значение разделителя '@'.</p>
максимальная длина слова	0<=число<=3 2	<p>Числовое значение, ограничивающее длину слова, помещаемого в словарь. Длинные слова усекаются до максимально разрешенной длины.</p>
минимальная длина слова	0<=число<=3 2	<p>Числовое значение, ограничивающее длину слова, помещаемого в словарь. Слова, имеющие длину менее минимальной, в словарь не включаются.</p>

управление выводом	строка до 19 символов	Строка команд, управляющих размещением материала при отображении документов на экране, распечатке или выводе в файл для последующей обработки. Команды и, если необходимо, их числовые параметры записываются подряд без каких-либо разделителей:
Формат	Fчисло	задает горизонтальный формат размещения текста поля в символах;
Горизонтальная позиция	Xчисло	задает горизонтальную позицию, с которой начинается размещение материала поля;
Вертикальная позиция	Yчисло	задает позицию первой строки материала поля;
С новой строки	N	вызывает размещение материала поля с новой строки. Размещение материала полей, для которых не задано команды "новая строка", начинается со строки, в которой начинался материал последнего поля, имеющего команду 'N' в списке команд управления выводом;
Непосредственное следование	G	задает размещение материала поля непосредственно после материала предшествующего поля;
Абзац	Aчисло	задает величину абзацного отступа в символах;
Обратный отступ	Iчисло	задает набор с обратным отступом на величину числа в символах;
Центрирование	C	центрирует текст поля по формату;
Выравнивание влево	L	выравнивает текст поля влево по формату;
Выравнивание вправо	R	выравнивает текст поля вправо по формату;
Повторение заголовка	Oчисло	текст заголовка поля выводится подряд указанное число раз;
Группа	Nчисло	специфицирует для группы повторяющихся полей (т.е. представленных в БД несколькими сегментами - подполями, имеющими одинаковую метку) вывод материала подполями в порядке определения в схеме. В группу объединяются поля одного уровня, задаваемого параметром команды, и расположенные подряд. После спецификации команды схема должна быть перезагружена;
Сортировка	Uчисло Dчисло 0<число<7	специфицируют поля, значения которых будут использоваться в качестве ключей сортировки, соответственно по возрастанию (U) или убыванию (D) для упорядочения множества найденных документов по значению этого поля. Число указывает приоритет ключа;
Исключающая сортировка	U8 D8	специфицируют поля, значения которых будут использоваться в качестве исключаящих ключей

Вывод пустого поля	Ечисло	<p>сортировки, соответственно по возрастанию (U) или убыванию (D) для упорядочения множества найденных документов с исключением из результата всех документов, кроме первого, имеющих одинаковое значение этого поля.</p> <p><число>=8 указывает исключаящий ключ.</p> <p>После спецификации команды схема должна быть перезагружена;</p> <p>специфицирует для пустого поля вывод на экран заголовка и специального знака (код '08'). Параметр задает вертикальный размер окна для ввода значения поля и используется обычно при редактировании полей документа;</p>
Подавление вывода	'.'	<p>подавляет вывод поля. Должна находиться в первой позиции строки команд;</p>
Избирательный вывод сегментов	'*' *число	<p>специфицирует вывод только тех подполей поля, по которым документ был отобран при поиске;</p> <p>специфицирует вывод подполя, порядковый номер которого задан числом;</p>
Вывод разделителей	'#' '!'	<p>специфицирует вывод списка разделителей предложений перед выводом поля. Если символ '#' приведен в определении поля-начала документа, то соответствующий список разделителей предложений будет выведен перед началом документа;</p> <p>специфицирует вывод списка разделителей предложений после вывода поля. Если символ '!' приведен в определении поля-начала документа, то соответствующий список разделителей предложений будет выведен после конца документа;</p>
Подавление вывода	S '-' -число ' ' ' '	<p>подавляет вывод подполя, имеющего пробельное значение;</p> <p>для поля-заполнителя подавляет его вывод, если перед ним выводился заполнитель с командой "-";</p> <p>подавляет вывод информационного поля, если поле с указанной меткой (число) имеет в документе не пустое значение;</p> <p>отменяет подавление вывода условного заполнителя (FILLER, TRAILER, значение метки поля которого указывает на ведущее поле) в последнем экземпляре группы или (в режиме вывода в IRBIS-формате) специфицирует вывод поля со значением метки, приведенной в заголовке этого поля.</p>
Файл специальной обработки	имя файла	<p>специфицирует имя текстового файла, находящегося в текущей директории.</p> <p>Файл с расширением 'STP' содержит слова, которые не должны включаться в словарь поля.</p> <p>Файл с расширением 'PLS' содержит слова, которые должны включаться в словарь поля. Такие слова записываются в файл отдельными текстовыми строками и должны быть упорядочены по кодам алфавита.</p> <p>Файл удаляемых символов имеет расширение 'PRG'. При загрузке документов символы (байты), приведенные в единственной строке файла, будут удаляться из этого поля.</p>

Практическая работа 3. Сравнительный анализ ИПС

1. Работа заключается в сравнительном изучении систем Апорт, Яндекс, Рамблер. Результаты изучения студент должен отразить в виде таблицы в файле Отчет_ФИО (ориентация таблицы — альбомная). В ячейках записать, как в каждой системе представляется тот или другой элемент языка запросов или входного/выходного интерфейса (все допустимые способы). В некоторых случаях можно отвечать знаками «+» или «-» (например, «*Описание документа*») или свободным текстом своими словами (например, «*Релевантные страницы того же сайта*» или «*Сортировка*»).

2. Перейти на сайт поисковой системы Апорт (затем Яндекс и Рамблер). Найти в каждой системе ссылки на ее описание в целом, на описание языка запросов, интерфейсов («*Справка*», «*Помощь*», «*Расширенный поиск*» и т.п.). Перейдя по ссылкам, внимательно изучить справочную информацию и в рабочей тетради кратко законспектировать основные пункты. После этого для каждой системы заполнить соответствующие ячейки таблицы (разделы 1, 2).

Примечание. Если текст ответа не помещается в ячейке таблицы, рекомендуется делать сноску и продолжать его под таблицей. Обратить внимание на то, что возможности систем в простом и в расширенном поиске различаются. Отразить это в отчете. Обратить внимание на наличие разделов «*другое*».

3. Вернуться назад на начальную страницу поисковой системы Апорт (затем Яндекс и Рамблер). Ввести какой-либо запрос (например, «*Статистические методы в лингвистике*») в окне для текстового запроса и провести поиск. Страницу с результатами поиска сохранить в своей папке в формате «*только html*».

4. Изучить форму представления результатов. Кратко записать в тетради, что содержится на веб-странице с результатами поиска (структуру веб-страницы). Изучить форму представления отдельных веб-документов (их краткие описания с дополнительной информацией). На основе изучения полученных результатов и ранее изученной справочной информации заполнить соответствующие ячейки таблицы (раздел 3).

5. Предъявить работу преподавателю.

Результаты сравнительного изучения систем Апорт, Яндекс, Рамблер

<i>№ раздела</i>	<i>Параметры</i>	<i>Апорт</i>	<i>Яндекс</i>	<i>Рамблер</i>
	<i>Поиск по тексту</i>			
1.	<i>Логические операторы:</i>			
1.1	конъюнкция			
1.2	дизъюнкция			
1.3	отрицание			
	<i>Синтагматические операторы:</i>			
1.4	фразы (словосочетания, слова рядом)			
1.5	расстояние в словах			
1.6	расстояние в предложениях			
1.7	<i>Морфологическая нормализация (автоматическая, используемые метасимволы)</i>			

2.	Поиск по полям			
2.1	по заглавию			
2.2	по полю ключевых слов			
2.3	по комментарию к картинкам (поле ALT)			
2.4	по тексту гиперссылок			
2.5	по адресам ссылок			
2.6	по доменному имени сайта (сервера)			
2.7	по языку			
2.8	по дате			
2.9	по формату			
2.10	Другие			
3.	Интерфейс выдачи (форма представления результатов)			
3.1	статистика слов из запроса			
3.2	количество найденных документов			
3.3	количество найденных сайтов			
3.4	количество документов на странице результатов			
3.5	сортировка документов на странице выдачи			
3.6	поиск в найденном			
3.7	описание документа включает следующие элементы:			
a)	URL (адрес в сети)			
b)	размер документа (объем)			
c)	дата создания			
d)	кодировка			
e)	аннотация (краткое содержание)			
f)	указание на другие релевантные веб-страницы того же сайта			
g)	ссылка на рубрику каталога, к которой относится найденный документ или сайт			
h)	поиск похожих документов			
i)	другое			

Практическая работа 4. Разработка поисковых запросов в русскоязычных вербальных ИПС

Составление и отладка тематического запроса

1. Составить в тетради запрос по теме «Морские сражения во время Великой отечественной войны». При этом убрать из темы незначимые слова, расширить запрос синонимами, составить логическую формулу запроса с обязательным использованием операторов конъюнкции, дизъюнкции, расстояния и фразы (жесткое словосочетание).

2. Показать запрос преподавателю.

3. Затем записать его варианты на языках систем Апорт, Яндекс, Рамблер.

4. Отладить запрос в режиме реального поиска, проводя последовательно сеансы во всех трех системах. Попытаться варьировать поисковые предписания, чтобы добиться оптимальных показателей поиска. Для этого фиксировать в тетради полученные результаты по каждому варианту: точность (по первым 20 документам) и условную полноту (абсолютный объем выдачи).

5. Вернуться к наилучшему поисковому предписанию и текст запроса скопировать через буфер обмена из поисковой строки (окно для ввода запроса) в окно файла отчета Report3 (поочередно в каждой системе). Указать при этом в отчете показатели точности и полноты. Первую веб-страницу с результатами поиска в каждой системе сохранить в своей папке в формате «только html».

Знакомство с поиском по полям («Расширенный поиск»)

6. Найти с помощью системы Яндекс документы, посвященные Льву Гумилеву. Количество найденных документов и сайтов записать в файл отчета. Адрес (URL) первого документа из списка сохранить в Избранном в папке «Гумилев».

7. Затем перейти в режим расширенного поиска и найти документы, посвященные Льву Гумилеву, с датой после 1 октября 2004 г. Новое количество найденных документов и сайтов снова записать в файл отчета. Первый документ из списка результатов поиска сохранить в своей папке в формате «*веб-архив, один файл*» (*.mht).

8. Найти через систему Рамблер документы по теме «Экономика города Москвы». При этом объем выдачи (количество описаний документов на странице результатов) установить равным 30. Результаты поиска отсортировать по дате (по убыванию) и первую веб-страницу с результатами поиска сохранить в своей папке в формате «только html». Количество найденных документов и сайтов зафиксировать в файле отчета.

9. Перейти в режим расширенного поиска и найти документы по той же теме, но находящиеся лишь на сайте www.rbc.ru. Результаты поиска отсортировать по дате (по возрастанию) и первую веб-страницу с результатами поиска сохранить в своей папке в формате «только html». Количество найденных документов и сайтов зафиксировать в файле отчета.

10. Найти через систему Яндекс документы по теме «Образование», из которых есть ссылка на сайт www.rsuh.ru. Первую веб-страницу с результатами поиска сохранить в своей папке в формате «только html». Количество найденных документов и сайтов зафиксировать в файле отчета.

11. Загрузить один из найденных документов, просмотреть его html-код, найти в нем ссылку на сайт www.rsuh.ru и элемент гиперссылки (от начального до конечного тега А) через буфер обмена скопировать в файл отчета.

12. Документ в формате mht, сохраненный в п. 7 (о Льве Гумилеве), прочитать в редакторе Word: вначале в формате веб-страницы, затем в формате «только текст». При втором чтении просмотреть содержимое окна ввода редактора Word (особенно начало и конец файла), скопировать первую страницу окна ввода в файл отчета и быть готовым объяснить, что такое формат mht.

Примечание. Формат *mht* кодируется в соответствии со стандартом *MIME* (*RFC2046* и *RFC2047*).

13. Предъявить работу преподавателю.

Практическая работа 5. Сравнительный анализ глобальных ИПС вербального типа

1. Работа заключается в сравнительном изучении заданных сети Интернет.

Примечание. Набор систем и их количество может меняться по усмотрению преподавателя.

2. Перейти на сайт соответствующей поисковой системы (здесь и далее — доменное имя системы: *www.название_системы.com*). Найти в каждой системе ссылки на ее описание в целом, на описание языка запросов, интерфейсов, режимов работы и других особенностей системы. Описание каждой ИПС кратко законспектировать в тетради.

3. Проанализировать и сравнить возможности систем в режиме расширенного поиска. Страницы интерфейса расширенного поиска сохранить в своей папке.

4. Результаты анализа в сжатом виде представить в форме сводной таблицы (с. 38) в файле отчета Report4 (ориентация таблицы — альбомная). Размеры таблицы можно увеличить. Если что-то не помещается в таблице, в ячейке делать сноску на текст под таблицей (таблица не столько форма представления результатов, сколько схема анализа).

5. Предъявить работу преподавателю.

Результаты сравнительного изучения глобальных вербальных ИПС

№ п/п	Параметры	Wis e-nut	Goo -gle	Allt he Web	Teo ma
1	Логические операторы (какие и как задаются)				
2	Синтагматические операторы (какие и как задаются)				
3	Поиск по полям (составить список полей, отмечать их наличие / отсутствие в конкретных системах)				
	<i>поле 1</i>				
	<i>поле 2</i>				
				
	<i>поле k</i>				
4	Выбор поисковой базы данных (в каких ресурсах можно искать)				
	<i>ресурс 1</i>				
	<i>ресурс 2</i>				
				
	<i>ресурс k</i>				
5	Формат выдачи содержит				

	следующие элементы (под таблицей привести пример из каждой системы)				
	<i>элемент 1</i>				
	<i>элемент 2</i>				
				
	<i>элемент k</i>				
6	Специальные возможности или характерные особенности (описать для каждой системы)				

Практическая работа 6. Глобальные вербальные ИПС: изучение и поиск

1. Провести поиск по теме «Компьютерная лингвистика» в заданных глобальных ИПС (*набор систем и их количество может меняться по усмотрению преподавателя*). Поисковое предписание логически должно выглядеть следующим образом:

(computational V computing V computer) & linguistics.

Запрос задать по-английски дважды, как конъюнкцию и как устойчивое словосочетание (фраза), используя характерные для каждой системы способы выражения операторов (для незнакомых систем найти соответствующую справочную информацию). Первую веб-страницу с результатами каждого поиска сохранить в своей папке в виде *«только html»*. Количественные результаты отразить в таблице:

№ п/п	Название ИПС	Найдено документов/сайтов		1. Комментарий (если требуется)
		Запрос как конъюнкция	Запрос как устойчивое словосочетание	
1	AlltheWeb (Fast)			
2	AltaVista			
3	Google			
4	MSN Search			
5	Teoma			
6	WiseNut			
7	Yahoo			

2. Проанализировать и попытаться объяснить полученные объемы выдачи в каждой системе.

3. В трех системах (Яндекс, Рамблер, Google) провести поиск по теме «Компьютерная лингвистика». Поисковое предписание, соответствующее логической формуле **«(компьютерная V вычислительная) & лингвистика»** (дизъюнкция устойчивых словосочетаний «компьютерная лингвистика» и «вычислительная лингвистика»), задать по-русски, используя характерные для каждой системы способы выражения операторов. Обратит внимание на различия в морфологической нормализации в разных системах. Первую веб-страницу с результатами каждого поиска сохранить в своей папке в виде *«только html»*. Количественные результаты отразить в файле отчета в виде таблицы.

4. Провести поиск по той же теме, сузив область поиска заданием дополнительных терминов, позволяющих отобрать документы, где наиболее вероятны определения данного термина (учебник, курс лекций, определение, называется и т.п.). Количество найденных документов указать в файле отчета. Туда же через буфер обмена скопировать поисковые предписания по каждой системе.

5. Просмотреть в каждой системе найденные документы, найти в них по одному определению термина «компьютерная лингвистика» и записать их в файл отчета с указанием адреса (URL) соответствующего документа.

6. Предъявить работу преподавателю.

Практическая работа № 7 Глобальные вербальные ИПС: изучение и поиск

1. С помощью систем Yahoo, AltaVista, Google найти материалы по психологическим (psychological) аспектам взаимодействия человека и компьютера (human computer interaction), опубликованные на английском языке в 1998 году. При составлении поискового предписания обязательно использовать аббревиатуру HCI (human computer interaction). По каждой системе первую веб-страницу с результатами поиска и один из документов сохранить в своей папке в формате «только html». Количество найденных документов и текст поискового предписания в каждой системе сохранить в таблице в файле Report6.doc в своей папке. Цель — правильное составление запроса и знакомство с системами. Текст поискового предписания записать также в рабочей тетради.

2. Найти в системе AltaVista фотографии с изображениями Ниагарского водопада (Niagara falls) (поиск в БД изображений — images). Одну из них сохранить как графический файл в своей папке на диске, а ее URL — в своей папке в Избранном.

3. Провести поиск на английском языке в системе Vivisimo по теме «Русская литература». Первую веб-страницу с результатами поиска скопировать в файл отчета.

4. С помощью системы AlltheWeb (Fast) выявить материалы на русском языке по автоматическому синтаксическому анализу. Цель — правильное составление поискового предписания. Наилучший вариант поискового предписания через буфер обмена скопировать в файл отчета. Первую веб-страницу с результатами поиска сохранить в своей папке в формате «только html».

5. Используя системы Teoma и WiseNut, найти веб-страницы на английском языке, посвященные изучению аномальных/ паранормальных явлений в России. Для перевода запроса на английский язык и подбора синонимов воспользоваться онлайн-словарями (www.multitran.ru, Lingvo на сайте Яндекса и др.). Отметить особенности языка запросов в каждой системе. Веб-страницу с результатами поиска по каждой системе сохранить на диске в своей папке в формате «только html».

6. Предъявить работу преподавателю.

9.2. Методические рекомендации по подготовке письменных работ.

Письменными работами по данной дисциплине являются рефераты, а также отчеты о практических работах, которые обучающиеся выполняют и оформляют в соответствии с методическими рекомендациями по подготовке письменных работ.

9.2 Методические рекомендации по подготовке письменных работ, требования к их содержанию и оформлению

Порядок составления и оформления отчета о практической работе

В значительной мере эффективность решения задачи по выполнению практической работы зависит от качества соответствующего отчета. Для этого необходимо соблюдать следующие основные требования по составлению и оформлению отчета, обусловленные соответствующими нормативными документами. Текст отчета должен быть лаконичным и

вместе с тем информативным. Текст должен быть изложен с соблюдением правил грамматики. Отчет составляется с обязательным составлением следующих разделов:

1. Заголовок отчета.
2. Цели работы.
3. Методика работы.
4. Порядок выполнения работы (этапы работы).
5. Выводы по работе.

1. В **заголовке отчета** приводятся наименования идентифицирующих признаков: **Отчет о практической работе № 1** по теме, например, «*Описание документов средствами различных ИПЯ*», ниже указываются данные студента (фамилия и инициалы, вид обучения, специальность, курс, группа).

2. В разделе **Цель работы** формулируется цели работы студента в соответствии с содержанием раздела «Постановка задачи» данной работы и индивидуального задания студенту на работу.

3. В разделе **Методика работы** указывается методика работы в соответствии с имеющейся формулировкой в разделе «Методика работы» данной работы и при необходимости уточняется в зависимости от содержания конкретного варианта задания студенту на практическую работу.

4. **Порядок выполнения работы.** Приводятся номера и наименования этапов работы, предусмотренные для работы данного Практикума. По каждому из этапов приводится описание выполненных студентом работ, направленных на достижение цели работы. Пропуск какого-либо из этапов работы Практикума не допускается. В рамках этапов помещается соответствующий иллюстративный материал - таблицы, рисунки (графики), полученные по ходу решения задачи работы. Обозначение иллюстративного материала выполняется в соответствии с правилами, принятыми для публикаций. Обозначение каждой таблицы и рисунка должно иметь номер и наименование. Внутри каждого отчета таблицы и рисунки обозначаются соответственно сквозными номерами. Обозначение таблицы указывается над таблицей, а обозначение рисунка под рисунком. Приводимые в тексте данной работы примеры включать в отчет не разрешается. Применяется только материал, полученный в ходе работы студентом по соответствующему заданию, полученному от преподавателя.

5. Последним разделом отчета являются **выводы** по работе. Это самая сложная и трудная часть работы. Очень важно, чтобы выводы отражали методику, технологию, применяемые программно-аппаратные средства решения задачи. Полезно каждому из этапов работы формулировать не менее одного вывода. Вывод может содержать от одного до трех предложений. Формулировки выводов должны быть конкретными, информативными, лаконичными, по возможности подкрепляться количественными данными.

Оформление отчета выполняется с учетом общепринятых правил. Графическая часть отчетов должна соответствовать правилам графического оформления. Текст отчета набирается в редакторе Word через 1,5 интервала, 14 кегль. Следует использовать шрифт Times New Roman. Заголовки разделов и подразделов выделяются жирным шрифтом. После окончания оформления отчета он проверяется студентом на предмет качество содержания и формы. При условии обнаружения ошибок последние исправляются. После устранения дефектов отчета его экранная форма, или принтерная распечатка предъявляется преподавателю. При условии обнаружения преподавателем ошибок в отчете студент их исправляет и предъявляет отчет преподавателю повторно. Если ошибок нет, то отчет принимается и сохраняется на жестком диске.

Отчет по работе сохраняется студентом в виде отдельного файла. В имени файла указывается фамилия студента и номер выполненной работы. Файл сохраняется в папке с фамилией студента в папке соответствующей студенческой группы. Папка группы создается на первом занятии. В имени папки группы должен присутствовать индекс группы. Папка группы включается в папку «Мои документы».

АННОТАЦИЯ РАБОЧЕЙ ПРОГРАММЫ ДИСЦИПЛИНЫ

Целью освоения дисциплины «Проектирование поисковых машин в Интернет» является теоретическое и практическое освоение методов и технологий создания, формирования и ведения современных баз данных с использованием информационно - поисковых систем, их технологий и языковых средств, а также особенностей поиска информации с их помощью и аналогичными поисковыми возможностями поисковых машин Интернет, с учётом особенностей поиска информации с их помощью и применение этих навыков в различных областях деятельности.

Задачи дисциплины: раскрыть эволюцию исследований в области информационного поиска; охарактеризовать понятие информационного поиска и классификацию его видов; рассмотреть методические основы информационного поиска; охарактеризовать особенности и средства поиска информации в глобальных сетях.

Знать: особенности информационно – поисковых систем (ИПС), как мощного программного средства создания, ведения баз данных и организации поиска в них данных, языковые средства поиска информации с использованием ИПС и поисковых машин, различия между СУБД, ИПС и поисковыми машинами, технологии применяемые в ИПС для создания, ведения баз данных и обслуживания абонентов на их основе, знать основные современные программные изделия класса ИПС и основные поисковые машины, знать основные современные программные изделия класса ИПС; методы информационного обслуживания; назначение и виды ИКТ; технологии сбора, накопления, обработки, передачи и распространения информации их помощь.

Уметь: проектировать и создавать базы данных и управляющие ими приложения на основе ИПС, организовывать ведение словарей в ИПС, применять языковые средства ИПС и поисковых машин для поиска информации в информационных массивах оценивать качество и затраты проекта; разрабатывать технологическую документацию; использовать необходимые функциональные и технологические стандарты; использовать документальные информационные системы и поисковые машины для информационного обслуживания, а также для проектирования ИС и Архивов, организовывать ведение словарей в ИПС, применять языковые средства ИПС и поисковых машин для поиска информации в информационных массивах, создавать пользовательскую документацию.

Владеть: навыками работы с инструментальными средствами ИПС, средствами поиска информации в диалоговом и пакетном режимах, в том числе с использованием постоянно – действующих запросов, навыками формирования формы экранов для обслуживания пользователей на основе баз данных ИПС, способностью организовывать индексирование информационных массивов на ПЭВМ с использованием поисковых машин; навыками разработки технологической документации и использования необходимых функциональных и технологических стандартов.